DOCUMENT RESUME

ED 200 619 TM 810 184

AUTHOR Michener, R. Dean; And Others
TITLE Automating Exams for a Statistics Course: II. A Case Study.
PUB DATE May 78
NOTE 40p.; Paper presented at the Annual Convention of the Association for Educational Data Systems (16th, Atlanta, GA, May 16-19, 1978).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Computer Programs; Higher Education; *Item Banks; *Statistics; *Test Construction
IDENTIFIERS *Statistical Test Item Collection System

ABSTRACT

A specific application of the process of automating exams for any introductory statistics course is described. The process of automating exams was accomplished by using the Statistical Test Item Collection System (STICS). This system was first used to select a set of questions based on course requirements established in advance; afterward, STICS was utilized to generate multiple exams from this collection of questions. The procedure for automating exams involved five steps: (1) establishing course objectives for the material to be covered; (2) selecting a question frame, or collection, to reflect the established objectives; (3) editing and extensive review of the questions considered for use; (4) refining the question frame and possible additions to it; and (5) generating exams using the conversational program BEGET. These five steps were covered in detail. The exam production process was illustrated by applying STICS in the context of an imaginary course that purported to promote student adoption of the critical, questioning outlook expressed in Huff's book, How to Lie with Statistics. Fifteen objectives defining what is expected of students completing the course, and exam questions and answers are provided in appendices. (Author/RL)

AUTOMATING EXAMS FOR A STATISTICS COURSE

II. A CASE STUDY

R. Dean Michener, Alvin Bugbee, and Jerry A. Warren
University of New Hampshire

This paper describes a specific application of the process of automating exams for a course. The process of automating exams was accomplished by using the Statistical Test Item Collection System, STICS (J. Warren, et al., 1978). This system was first used to select a set of questions based on course requirements established in advance; afterward, STICS was utilized to generate multiple exams from this collection of questions.

Since our purpose is to show how STICS can be used to prepare exams for any introductory statistics course, we chose to illustrate the exam production process by making up a new course rather than use one of the courses involved in the development of STICS. The book How to Lie with Statistics (D. Huff, Norton & Co., N.Y., 1954) was chosen as a starting point for our new course. The application of STICS is in the context of an imaginary course that purports to promote student adoption of the critical, questioning outlook expressed in How to Lie with Statistics.

THE PROCEDURE

The procedure for automating exams involves five steps: (1) establishing course objectives for the material to be covered; (2) the selection of a question frame, or collection, to reflect the established objectives; (3) editing and extensive review of the questions considered for use; (4) further refine-

ments to the question frame and possible additions to it; and (5) generating exams. These five steps will be covered in detail.

Step 1 was to establish our objectives for this new course. Guidelines in Preparing Instructional Objectives (R.F. Mager, Fearon, Belmont, Calif., 1962) were followed to determine our course objectives except that we did not attempt to follow Mager's approach for the establishment of specific minimum requirements. We found a thorough re-examination of the material in How to Lie with Statistics necessary in order to identify explicitly what behavior, or abilities, would be required of students when they finish a course on this material. For example, Huff devoted an entire chapter to the three common "averages": mean, median, and mode. These three measures were clearly distinguished with examples of which types of situations would best have the "average" defined as a mean, median, or mode. Applications as well as computational examples were provided. We agreed on the following three specific behavioral objectives to cover this topic:

(1) Given a value said to be an average, a student should recognize the need to know whether the average is a mean, median, or mode.
(2) A student should be able to identify which average is most applicable, or informative, for a given situation.
(3) Given a data set, the student should be able to determine the mean, median, and mode.

A total of 15 objectives were decided upon to initially define what is expected of students completing this course (see Appendix 1). While we recognized that this set of objectives should be regarded as a trial set, we also hoped

that it would not be necessary to return to this activity.

This process of establishing course objectives for <u>How to Lie with Statistics</u> proved time consuming. Identifying and stating succintly what we expected of student behavior after completing the material took approximately three days. (This time investment was made up of one to three hour sessions at the ends of several days rather than three consecutive days.)

<u>Step 2</u> involved the selection of a question frame (collection) based on the objectives established in Step 1. This required (a) choosing keyterms likely to be associated with questions related to our choice of objectives; and (b) running the STICS program SELECT to extract these questions from a large collection of questions and answers, the National File.

Each item in the National File has associated with it a series of keyterms and question characteristics (Technical Report). Generally, keyterms indicate statistical topic coverage while question characteristics describe such things as estimated difficulty, question type (essay, multiple, choice, ...) etc. Our plan was to initially select on the basis of keyterms only in hopes of forming a large pool of questions that could later be refined using question characteristics.

We began by searching the list of keyterms for the National File (see Appendix 2) to find terms likely to be associated with our own set of objectives. For example:

<u>Example 1</u>

Objective 6. Given a data set, a student should be able to determine the mean, median, and mode.

Keyterms chosen: MEAN,MEDIAN,MODE

Example 2

Objective 7. A student should be able to identify graphs or charts which have been incorrectly labeled or have truncated axes.

Keyterms chosen: TYPICALPIC/GRAPH, COMMONPITFALLS, HISTOGRAM

Example 3

Objective 3. A student should be able to define and explain the advantages of random sampling and stratified sampling.

Keyterms chosen: SIMPLERANDOM,STRATIFIED

We spent around two hours choosing keyterms.

Next, we used the conversational program SELECT several times to obtain a collection of questions linked to our choices of keyterms. SELECT was used more than once partly because of a system crash and partly because we committed a classical blunder. (SELECT begins by asking if the user is an expert. We modestly said yes. And missed a message about a number of changes that had occurred since our last use of the program.) Following these difficulties, which stretched over a two day period, we did make a 45 minute run to extract an initial question pool of 194 questions, FRAME(1). We expected that many of these items would have to be discarded, but hoped that it would provide most of the items needed for preparing exams for our new course. A sample run of SELECT appears in Appendix 3.

Step 3 involved the examination and subsequent editing of FRAME(1). The full set of questions were printed so we could review each item. We decided that 53 questions (27%) were acceptable including those that needed minor modi-

fication. Since we wanted a larger item pool, we reconsidered our initial run of SELECT and made some changes. The usual search basis used by SELECT is to retrieve items on the basis of what are called primary keyterms. These are terms that the question calssifier regards as closest to the central topic(s) of a question. In addition, secondary keyterms also are assigned to items. For example, a keyterm such as COMMONPITFALLS isn't likely to be attached to a question as a primary keyterm, but it often will appear as a secondary keyterm. We modified our use of SELECT so that it would search for both primary and secondary keyterms and made use of additional keyterms that we had neglected previously. Our new run of SELECT yielded 43 new items. We found only eight of these (19%) acceptable. The total acceptable now was 61.

At this point we decided to reconsider our initial set of objectives. We concluded that we had allowed How to Lie with Statistics to dominate our choice of objectives to the point that we had neglected to include any objectives that could not be supported by that book alone. We had objectives that faithfully reflected How to Lie with Statistics, but not our own judgements of what students should gain from our course. We then revised our set of objectives, largely by introducing objectives that include computational activities and the notion of confidence intervals (objectives 13 through 17 in Appendix 4).

When How to Lie with Statistics was published (1954) it was probably very realistic to do no more than mention that a standard error is a useful piece of information. But, now it seems unwise not to have students learn that they themselves can obtain this useful piece of information (using a pocket calculator or computer) and to include doing so among the objectives of this kind of course. Similar arguements led to the addition of new objectives such as those dealing with confidence limits for a mean, etc.

Again we returned to SELECT and again we extracted and refined a new pool of questions. This reconsideration of objectives and re-extraction of items took three days to provide us with a collection of 78 items.

Step 4 involved extensive editing of our question frame to prepare for exam generation. In doing this we set up our question collection so that exams could be prepared by using random sampling within two question groups — one consisting of questions requiring less than five minutes, the other consisting of questions rated five minutes or more. Within each of these groups we chose to have the frequency of questions dealing with an objective dependent on the relative importance of the objective. Question frequencies were in a 3:2:1 ratio with the highest frequency applying to the most important objectives. Table 1 summarized the distribution of questions that we decided upon.

In arriving at this distribution of questions we found it necessary to discard 31 questions and to write 17 new questions. If we had not made this adjustment we would not have been in a position to employ the simple random sampling technique that later was used in forming exams. We would have had a great excess of questions involving computational skills and few, if any, questions concerned with several other objectives. This process of adjustment required approximately four days of effort. Most of this time was consumed in writing new items and adding them to the question frame. Although we had hoped for even more, one of our rewards for having used the National File was that 47 items did not have to be written.

Step 5 was actually generating a series of exams using the conversational program BEGET. We spend around an hour using a terminal to obtain the three exams listed in Appendix 5 (answer guidelines were also obtained). Since we

TABLE 1 — Distribution of questions in the frame sampled for preparing exams

OBJECTIVE	RELATIVE IMPORTANCE *	QUESTIONS IN FRAME **
1. Identify how sample information was obtained	M	4
2. Identify bias in a sample	M	4
3. Define and explain random and stratified samples	H	6
4. Distinguish among the three common averages	M	4
5. Identify which average is most applicable	L	2
6. Calculate a mean, median, and mode	H	6
7. Identify misleading or poorly constructed graphs	M	4
8. The effect of sample size on results	M	4
9. Define and explain level of significance	M	4
10. Define and explain the standard error of a mean	L	2
11. Recognize use of standard and/or probable error	L	2
12. Recognize that small differences may be meaningless	L	2
13. Use of a random numbers table	L	2
14. Recognize the interrelationship of significance level, variability, tolerable error, and sample size	H	6
15. Calculate a range and standard deviation	M	4
16. Test for a significant difference between means	L	2
17. Create a confidence interval for the mean	H	6
	TOTAL QUESTIONS	64

*H: High importance; M: medium importance; L: Low importance

**Half of these questions belong to each of the two groups based on estimated time time required

used a special ribbon and forms in printing these exams we also had an overnight delay before they were available. Usually we can print exams within 30 minutes of running BEGET.

BEGET makes provision for generating exams by sampling groups that can be defined using either keyterms of question characteristics. Keyterms used can be either those that appear in the National File or new terms added during the editing process. In this case, our question frame had been set up so that we didn't need to use keyterms at all when exams were generated. We only needed to specify time required in answering questions to form these exams. For each exam we asked for one random sample of four questions from the group of questions requiring less than five minutes and a second random sample of eight questions from the group requiring five minutes or more.

We would like to report that we were delighted with the three exams in Appendix 5 but feel obliged to point out that there were some difficulties. Answers were not provided for two questions in Exam 2. (This is being corrected as questions and answers in the National File undergo external review.) More importantly, Question 10 in Exam 3 was totally unacceptable. Somehow, a question asking for flaws in a graph that hadn't been printed penetrated our screening and editing activities and managed to come out of hiding in our first trial of the exam generator. Apparently, BEGET provides no protection against improper items. We humans are still needed.

DISCUSSION

In retrospect, these efforts to automate exams for our <u>How to Lie with Statistics</u> probably were more typical than we really wanted. We invested around ten days in setting up so that new exams can be generated with rela-

tively little additional investment of effort. While part of this start-up cost can be assigned to shortcomings in the files and programs now making up STICS, most of this time was required for activities that depend on people — including reading the instructions provided in "self-documented" programs such as SELECT.

Setting objectives takes time. So does resetting objectives. Looking at items selected to reflect objectives takes time. Discarding unacceptable items takes time. And writing new items takes more time. While programs may later be developed to aid in these activities, these are now and probably always will be predominantly people dominated activities.

Then where do schemes like STICS fit in? Even though the components of STICS used in this exercise are still evolving, they probably are adequate for picturing where the payoffs are. By and large, they seem to be as much attitudinal as they are time-saving. They are attitudinal in the sense that they make it seem credible that two or more people can concentrate their efforts during a period of a month or so and reach an agreement on course objectives that are expressed in the exams that students later will take. All of this can be done without a scheme like STICS, but it rarely is done. STICS makes it possible to convert Monday's agreements on objectives to Friday's discontent at the kinds of questions that those objectives lead to. It makes it possible to push this process fast enough so that arriving at an acceptable question frame doesn't require six exercises in recalling what's been done and where things stand now. It makes it realistic to believe that a course can be set up so that student exams consistently will be in agreement with course objectives over a period of time.

If a course can be brought past the start-up period so that exams are being produced, the time savings become reasonably clear. In 15 to 20 minutes a BEGET user now can set up five exams like those in Appendix 5 These are exams ready for reproduction and accompanied by answer guidelines. The savings in time are easy to see.

The problem is arriving at a state that will support exam generation. Here STICS offers help that will vary widely from case to case. If STICS supplies 3/4 of the questions and answers needed for a course, the time savings will be appreciable. But, our experience to date suggests that 3/4 may be high for the current version of the National File and that for some people it never will be high. Sometimes the reaction of people extracting questions is "I didn't like most of the questions, but I can look at them and write new ones that I do like." Then STICS saves time, but much less time than when hand crafting is regarded as necessary.

While the procedure we followed may seem lengthy and time consuming, two important features should be noted:

1) start-up efforts may be substantial, but these efforts should be repaid over several years to come in terms of clear course expectations and corresponding exams;

2) there should be improved consistency of course requirements.

The purpose of this paper is to make explicit a means of using the STICS system to clarify objectives for a course and create examinations based upon those objectives. We attempted to show one means of utilizing the system for an imaginary course using Huff's book How to Lie with Statistics. While our example kept to a well-defined small subset of material, the procedure

we followed would be the same for any course objectives under consideration. Decisions on objectives weightings, etc. may indeed differ, but we hope to have indicated a procedure which can benefit most people involved with introductory statistics instruction.

REFERENCES

Huff, D. (1954), How to Lie with Statistics, New York: W.W. Norton & Company, Inc.

Mager, R.F. (1962), Preparing Instructional Objectives, Belmont: Fearon Publishers.

Warren, J.A., Michener, R.D., Bugbee, A. (1978) Automating Exams for a Statistics Course I. The Beginning of a Case Study, paper presented at the 1978 AEDS Convention, Atlanta, Georgia.

Geeslin, W.E., Bugbee, A. (1977) Data Bank Procedures for the Diagnostic System; Technical Report No. 1 for National Science Foundation Project No.: SED 76-12191, Available from Office of Academic Computing, University of New Hampshire, Durham, New Hampshire 03824.

Chapter 1

1. Student should be able to look for and identify how sample information was obtained for a given study.

2. Student should be able to identify bias in a sample as non-response, an unrepresentative sample, interviewer contamination, and tendency to give socially acceptable responses.

3. The student should be able to define the terms random sample and stratified sample and explain the advantages of each.

Chapter 2

1. Given an average, a student should be able to distinguish which among the three common averages (mean, median, mode) has been used or demonstrate some skepticism on the use of the average.

2. Student should be able to identify which common measure of an average is most applicable, or informative, for a given situation

3. Given a data set, the student should be able to determine the mean, median, and mode.

Chapter 3

1. Student should be able to identify graphs/charts which have been incorrectly labeled (or not labeled) or where axes have been truncated.

2. Given a statistical report, a student should be able to state the effect that a small sample size may have on the results.

3. Student should be able to explain why stating the level of significance is an important part of a report.

Chapter 4

1. Given a probable error or standard error, the student should explain or demonstrate how they are useful in interpreting a statistical report.

2. Given a comparison between figures with a small difference, the student should utilize the concept of standard error and recognize that small differences may be meaningless.

Extra

1. Student should be able to distinguish between the concepts of association and causality.

2. The student should be able to identify when a one-dimensional relationship has been misrepresented by an improper two-dimensional graph.

3. Student should be able to identify or give variables that are positively correlated or negatively correlated.

4. Student should be able to demonstrate that they can critically analyze a statistical report using the following five questions as guidelines.

    1. who says so?
    2. how does he know?
    3. what is missing?
    4. did somebody change the subject?
    5. does it make sense?

Tree Structure Listing of Keyterms

- PROBABILITY # 1
  - BASICTERMS/P # 2
  - PROBMODELS # 3
  - PROBDISTRIBUTION # 4
    - MULTIVARIATE/P # 5
    - UNIFORM # 6
    - JOINTDISTRIBUTIO # 7
    - TDISTRIBUTION # 8
    - FDISTRIBUTION # 9
    - XSQRDISTRIBUTION # 10
    - STANDUNITS/NORMA # 11
      - TSCORE # 12
      - ZSCORE # 13
      - OTHER/ST # 14
    - POISSON # 15
    - BINOMIAL # 16
    - NORMAL # 17
      - NORMALPAPER # 18
      - GRADINGCURVE # 19
      - OTHER/N # 20
    - MOMENTGENFN # 21
    - DENSITYFN # 22
    - CUMULATIVEFN # 23
    - CONCEPT/OTHER # 24
    - PROBFUNCTION # 220
    - GEOMETRIC # 228
    - HYPERGEOMETRIC # 229
  - EXPECTATION # 25
  - RANDOMVARIABLES # 26
    - RANDOMNUMBERS # 27
    - OTHER/RV # 28
    - DISCRETERANVAR # 190
    - CONTINUOUSRANVAR # 191
  - CONDITIONALPROBA # 29
    - NONBAYESIAN # 30
    - BAYESIAN # 31
  - COMBINATIONS # 32
  - PERMUTATIONS # 33
  - COUNTING # 34
  - SIMPLEPROBABILIT # 35
  - EVENTS # 36
    - OTHER/EVENTS # 37
    - INDEPENDENT # 38
    - DEPENDENT # 39
  - JOINTPROBABILITY # 214
- STATISTICS # 40
  - BASICTERMS/S # 41
  - NONPARAMETRIC # 42
    - CHISQUARE # 43
      - MCNEMARTEST # 193
      - GOODNESSFITTEST # 210
      - CONTINGENCYTABLE # 211
      - CONFIDLIMITVAR # 212
      - MEDIANTEST # 221
    - CORRELATION/NP # 44
      - RHO # 45
      - TAU # 46
    - DESCRSTAT/NP # 47

- MEDIAN # 48
- MODE # 49
- VARIABILITY/NP # 50
  - OTHER/V # 51
  - RANGE # 52
- SCATDIAGRAM/NP # 53
- BARCHART/NP # 54
- SKEWNESS/NP # 55
- POPULATIONMOD/NP # 56

OTHER/NP # 57
SIGNTEST # 222
WILCOXONSIGNRANK # 223

- CONCEPT # 58
  - DESIGN # 59
  - TESTING # 60
  - ESTIMATION # 61
    - SAMPDIST/C # 62
    - CONSISTENCY # 63
    - SUFFICIENCY # 64
    - EFFICIENCY # 65
    - UNBIASEDNESS # 66
    - CONFIDENCEINTERV # 67
      - SIMPLE/CI # 68
      - OTHER/CI # 69
    - MAXLIKELIHOOD # 70
    - MINVARIANCE # 71
    - ESTIMATION/OTHER # 72
  - POPULATION # 73
  - CENTRALLIMITTHM # 74
  - POWER # 75
  - TESTOFSIGNIFICAN # 76
  - TYPE1ERROR # 77
  - TYPE2ERROR # 78
  - TCHEBYSHEFFSTHM # 195
  - PREDICTION # 209
  - SCOPEOFINFERENCE # 224
- PARAMETRIC # 79
  - DESCRSTAT/P # 80
    - CENTRALTENDENCY # 81
    - SCATDIAGRAM/P # 82
    - PROPORTION # 83
    - BARCHART/P # 84
    - MOMENT # 85
    - MEAN # 86
    - VARIABILITY/P # 87
    - STANDARDDEVIATIO # 88
    - VARIANCE # 89
    - COVARIANCE # 90
    - STANDERROROFMEAN # 91
    - POPULATIONMODELS # 92
    - HISTOGRAM # 93
    - FREQDIAGRAM # 94
    - PERCENTILE # 95
    - FREQTABLE # 96
    - CLASSINTERVAL # 97
    - SKEWNESS/P # 98
    - STANDERROR/OTHER # 192
    - STEMLEAFPLOT # 194
    - VARIANCE/OTHER # 204
    - GRAPH/PICTOGRAPH # 227

- COEFFVARIATION # 230

- TTEST # 99
  - ONETAIL/T # 100
  - TWOTAIL/T # 101
  - OTHER/T # 102
- REGRESSION # 103
  - MULTIPLE/REG # 104
  - SIMPLE/REG # 105
  - RESIDUALS/REG # 106
  - STANDARDERROR # 107
  - MULTIVARIATE/REG # 108
  - OTHER/REG # 109
  - BASICTERMS/REG # 213
  - POLYNOMIAL/REG # 215
- ANCOVA # 110
- CORRELATION/P # 111
  - SIGNIFICANCE # 112
  - PARTIAL # 113
  - MULTIPLE/COR # 114
  - SIMPLE/COR # 115
  - BISERIAL # 116
- ANOVA # 117
  - OTHER/AN # 118
  - RESIDUALS/AN # 119
  - DEGREESOFFREEDOM # 120
  - COMPUTFORMULA # 121
  - FTEST # 122
  - TREATMENTASSIGN # 123
  - LATINSQUARE # 124
    - TWOWAY/LS # 125
    - ONEWAY/LS # 126
    - NWAY/LS # 127
  - RANDOMIZEDBLOCK # 128
    - TWOWAY/RB # 129
    - ONEWAY/RB # 130
    - NWAY/RB # 131
  - COMPLETELYRANDOM # 132
    - TWOWAY/CR # 133
    - ONEWAY/CR # 134
    - NWAY/CR # 135
  - MULTIVARIATE/AN # 136
  - SUMSOFSQUARES # 216
  - EXPERDESIGN/TERM # 217
- MULTIPLECOMPARIS # 137
  - APRIORI # 138
    - TRATIO # 139
    - FRATIO # 140
    - DUNNS # 141
  - APOSTERIORI # 142
    - LSD # 143
    - TUKEYSHSD # 144
    - SCHEFFESSMETHOD # 145
    - NEWMANKEULS # 146
    - DUNCANS # 147
    - DUNNETTS # 148
- ITEMANALYSIS # 149
- ZTEST # 205
  - ONETAIL/Z # 206
  - TWOTAIL/Z # 207
  - OTHER/Z # 208

SAMPLING # 150
PROBABILITYSAMPL # 151
PROBSAMPCONCEPT # 152
SIMPLERANDOM # 153
STRATIFIED # 154
CLUSTER # 155
SEQUENTIAL # 156
SAMPLINGDISTRIB # 157
SAMPLESIZE # 158
SAMPLE # 159
CENSUS # 160
FRAME # 161
SAMPLINGERROR # 162
PRECISION # 163
UNCORRELATED # 164
RANDOMVARIATION # 165
RANDOMNUMBERS/S # 166
CONVENIENCE/CHUN # 167
JUDGEMENTSAMPLIN # 168
NONSAMPLINGERROR # 169
ACCURACY # 170
OTHER/S # 171
NONRESPONSE # 172
STRUCTURALLIMIT # 173
OPERATIONALFAULT # 174
TESTTHEORY # 175
RELIABILITY # 176
VALIDITY # 177
TESTCONSTRUCTION # 178
MISCELLANEOUS # 226
MODEL # 196
ASSUMPTCUSTOMARY # 197
APPLICATIONEX # 198
TYPICALPIC/GRAPH # 199
SIMPLEDATASET # 200
TYPICALSUMMARY # 201
COMMONPITFALLS # 202
SPECIALCASE/PROB # 203
COMPUTERPROGRAMS # 218
UNH/ONLY # 219
I650/TEMPORARY # 225
I650I # 179
I650A # 180
I650C # 181
I650E # 182
I650F # 183
I650G # 184
I650H # 185
I650J # 186
I650K # 187
I650M # 188
I650N # 189

RUN SELECT

Are you a SELECT expert? NO

This is the Question Frame Extracting program. It allows you to specify keyterms or characteristics that you require of questions to be used on exams. You begin by specifying a single keyterm and the computer replies by indicating how many questions have that keyterm. You then may discard some of these questions by adding another keyterm or characteristic that you require (i.e. you now say that you only want questions having both the first and second requirement). The computer then reports how many questions satisfy your requirements. You may continue adding requirements or you may accept a group of questions. When you accept a group of questions, you define a segment. You may define several segments in a single run of this program. One or more segments comprise a question and answer frame. This program is the first step in storing a frame in a form, a file, that can be later sampled to make up exams.

During the run of SELECT, you will receive two kinds of prompt

Command: Which calls for a word response, YY>GCe response word HELP will result in a list of the words that can be used as a response and an explanation of what they do

Code No.: Which calls for a numeric response, e.g. -1 will result in a list of the values that can be used and an explanation of what they do

NEWSEG is used to provide an initial keyterm requirement for segment membership. KEYTERM allows specification of additional keyterm requirements, and CHARACTERISTIC allows specification of additional requirements based on characteristics such as Question Type, Cognitive Level, etc. As a general rule, all keyterm specification should be done before characteristics specifications is begun

When the questions were assigned keyterms, the person assigning them was asked - what keyterms would I use if I were only allowed a few terms for this question? The terms chosen are called primary keyterms. All searches are based on primary keyterms unless the command KEYSRCH is used to change this basis.

When this program selects for questions having certain characteristics, it will do so on the basis of questions having a characteristic "equal to" what has been specified. e.g. Find questions for which Time Required is equal to, say, 10 minutes. When this is not the basis that you want for dealing with characteristics, you may change to a different basis for a single search by entering -2 when Code No.: is requested for a particular kind of characteristic. Then you can search for questions on a basis such as not equal (NE), less than (LT), etc. Once the next search has been completed the search basis reverts to equality unless -2 is again used e.g.

```
Command:  CHARACT
Code No.:  5                (specifies Time Required)
Time Code No.:  -2          (change in relation)
Relational Command:  LE     (HELP produces a list of commands)
Time Code No.:  10          (Search is for questions with Time
                            Required less than or equal to 10 Minutes)
```

```
Are you using the National file? YES

There will be a short pause now as the Parent File is opened
Command: NEWSEG

    Keyterm Code No.: 161

There are now  2 Questions in Segment # 1
Command: DROP LAST

**** That Command is not valid at this time ****
Command: NEW S\S \SEG

    Keyterm Code No.: 152

There are now  11 Questions in Segment # 1
Command: SAVE

 11 questions are saved in Segment # 1
Command: NEWSEG

    Keyterm Code No.: 169

There are now  1 Questions in Segment # 2
Command: SAVE

  1 questions are saved in Segment # 2
Command: NEWSEG

    Keyterm Code No.: 172

There are now  1 Questions in Segment # 3
Command: SAVE
```

```
Command: NEWSEG

    Keyterm Code No.: 115

There are now  63 Questions in Segment #15
Command: SAVE

 63 questions are saved in Segment #15
Command: FINISH

Enter file name for storing your results
(Form: XXXXX where X is a letter): HUFF

Do you want your Frame order randomized?  NO


The next step in the process for creating your own
local file is to type the command
.DO CREATE
after you leave this program and your local file will be
created. If you simply want a printout of your file,
you should type
.DO PRINT
STOP

END OF EXECUTION
   TIME: 2.11  ELAPSED TIME: 16:50.02
```

Appendix 4

Revised Set of Objectives

Wt.		
2	1.	Student should be able to look for and identify how sample information was obtained for a given study.
2	2.	Student should be able to identify bias in a sample as caused by one of the following: non-response, an unrepresentative sample, interviewer contamination, or tendency to give socially acceptable responses.
3	3.	The student should be able to define the terms random sample and stratified sample and explain the advantages of each.
2	4.	Given an average, a student should be able to distinguish which among the three common averages (mean, median, mode) has been used or demonstrate some skepticism on the use of the average.
1	5.	Student should be able to identify which common measure of an average is most applicable, or informative, for a given situation.
3	6.	Given a data set, the student should be able to determine the mean, median, and mode.
2	7.	Student should be able to identify misleading or poorly constructed graphs/charts.
2	8.	Given a statistical report, a student should be able to state the effect that a small sample size may have on the results.
2	9.	Student should be able to define the term level of significance and explain why it is an important part of a report.
1	10.	Student should be able to define or explain the term standard error of a mean.
1	11.	Given a probable error or standard error, the student should explain or demonstrate how they are useful in interpreting a statistical report.
1	12.	Given a comparison between results with a small difference, the student should utilize the concept of standard error to recognize that small differences may be meaningless.
1	13.	The student should be able to explain the use of a random numbers table to generate a random sample.
3	14.	When using a simple random sampling procedure for estimating the mean or proportion of a population, a student should be able to utilize the inter-relationship between the following four quantities: level of significance, tolerable error, estimate of the variability of the population, and sample size.
2	15.	The student should be able to determine or interpret the range and standard deviation of a data set or distribution.

16. Given the means of two independent samples, the sample variances and the condition that the population variances are equal, the student should be able to test if the two means are significantly different.

17. Given a sample data set, the student should be able to calculate the standard error of the mean and use this to create a confidence interval for the mean.

How to Lie with Statistics
Spring, 1978
Exam - 1

Question # 1 5.0 Points
The range is a measure of
1. relative frequency
2. variability
3. central tendency
4. skewness
5. none of these.

Question # 2 5.0 Points
TRUE OR FALSE? IF FALSE, CORRECT IT.
IN A RANDOM SAMPLE WITH REPLACEMENT FROM A POPULATION, EACH INDIVIDUAL HAS THE SAME CHANCE OF BEING INCLUDED IN THE SAMPLE.

Question # 3 5.0 Points
The standard error of the mean is another name for the standard deviation of:

a. a sample
b. a population
c. the sampling distribution of any statistic
d. the sampling distribution of the mean
e. none of the above

Question # 4 5.0 Points
TRUE OR FALSE. IF FALSE EXPLAIN WHY.

If the results of an investigation show that one sleeping tablet works better than another at the 5% level of significance, the conclusion would be the same if tested at the 10% level of significance.

Question # 5 10.0 Points
You are the parent of two children that have been given an IQ test that purports to be a sampling of the intellect. One child receives a score of 99 and the other receives 101. You know, of course, that the IQ is based on 100 as average or 'normal'. Does this imply that one is "above average" and the other is "below average"? Explain.

Question # 6 10.0 Points

A report issued by a consulting firm retained by the operators of a large chemical plant is concerned with number of fish in a river before and after the plant was built. It states:

"Average number of fish before construction 100
Average number of fish one year after operation began 25

This difference is not sigificant at the .5% level, so there is no good reason to regard plant operation as harmful."

Do you agree with this conclusion? Explain.

Question # 7 10.0 Points

The results of a bar pressing experiment for a group of 9 rats yielded a mean equal to 25 presses/hour with a standard error of the mean equal to 5. If the population mean for bar pressing equals 10 presses/hour what is the probability of someone being able to repeat the experiment with results of a sample mean greater than 25?

Question # 8 10.0 Points

A newspaper editor wants to survey the attidudes of the public toward littering. Discuss the relative merits of using personal interviews or mailed questionnaires as methods of data collection.

Question # 9 10.0 Points

In the 1936 Presidential Election Franklin D. Roosevelt defeated Alfred E. Landon in a landslide vote. A Landon victory had been predicted by the Literary Digest, a magazine which ran the oldest, largest, and most widely publicized of the polls at the time. The Digest's final prediction was based on ten million sample ballots mailed to prospective voters and 2.3 million were returned. The sample of voters was drawn from lists of automobile and telephone owners. Despite the massive size of this sample, it failed to predict a Roosevelt victory, being off the mark by 19 percentage points. Explain why the Digest was so wrong.

Question # 10 10.0 Points

A SAMPLE IS CHOSEN BY NUMBERING ALL THE RED BOOKS IN THE LIBRARY AND THEN CHOOSING THE ONES THAT CORRESPOND TO RANDOM DIGITS IN A TABLE.
IS THE SAMPLE INDEPENDENT?
IS IT A SIMPLE RANDOM SAMPLE?
OF WHAT POPULATION?
(EXPLAIN YOUR ANSWERS REASONABLY)

Question # 11 10.0 Points

For the following data

8, 6, 9, 10, 6, 6, 11

Calculate:

a) mean

b) median

c) mode

d) range

e) variance

Question # 12 10.0 Points

IF THE OBSERVATIONS: 8, 11, 9, 17, 12, 15 ARE A SAMPLE OF SIZE 6 FROM N(MU, 24), THEN A 95% CONFIDENCE INTERVAL FOR MU HAS CONFIDENCE LIMITS

A) 12 +!- 3.92 B) 11.5 +!- 3.92 C) 12 +!- 5.142

D) 11.5 +!- 5.142 E) NONE OF THESE

Question # 1 5.0 Points
2. variability

Question # 2 5.0 Points
TRUE

Question # 3 5.0 Points
(d.)

Question # 4 5.0 Points
True, 95% confidence automatically implies more than 90% confidence.

Question # 5 10.0 Points
This question cannot be answered accurately without knowing the standard deviation for IQ scores. Then these differences from the mean or average can be evaluated relative to the standard deviation as to whether there is a significant difference between them.

Question # 6 10.0 Points
There are many reasons for disagreeing.

1. The Type 1 error rate level is very low (.5% = .005) which makes it more difficult to detect a difference.
2. We do not know how the sampling was carried out - was it random?
3. We do not know what size samples (or even if they were the same size) were used.

Question # 7 10.0 Points

n = 9
xbar = 25, MU = 10
SIGMA(xbar) = 5

Z = (25 - 10) / 5
= 3

Area beyond (Z = 3) = .0013

Prob. (someone obtaining an xbar > 25) = .0013

Question # 8 10.0 Points
Answer - One of the main problems with surveys, as in this particular case, is the response rate. The mail questionnaire is generally the least successful in terms of response rate. Personal interviews are generally much more effective, but do present problems with

respect to potential interviewer bias. Such problems as leading questions, questions which often produce socially acceptable responses, and even subjects' attempts to yield expected, or favored, responses often plague the design of an interview based survey.

The mail questionnaire can often be structured to avoid many of these difficulties, or sources of bias. However, if attempts are not made to increase the response rate, the sample may have a built-in bias. Various techniques used to increase response for mail surveys generally revolve around some effort to make the mailing appear personal. For example, using envelopes accurately addressed rather than simply " RESIDENT", providing stamped return envelopes, enclosing an inducement such as a pen or pencil, etc.

Question # 9   10.0 Points

The problem developed because the Digest relied on voluntary response and such samples are practically always biased. The respondents represented only that subset of the population with a relatively intense interest in the subject at hand and in no sense represents a random sample or representative sample.

Question # 10   10.0 Points

NONE PROVIDED

Question # 11   10.0 Points

a) mean = 8
b) median = 8
c) mode = 6
d) range = 5
e) variance = 26/(7 - 1) = 4.33

Question # 12   10.0 Points

(NOTE: TABLE IS REQUIRED)

A

This is a Cross-reference between the question numbers on this exam and the question numbers in the File

1-	37	2-	2	3-	29	4-	52	5-	28
6-	27	7-	53	8-	58	9-	6	10-	3
11-	40	12-	48						

Question # 1 5.0 Points

If the population mean of miles per week jogged by people over 65 equals 10 and the standard error of the mean for a group of 9 is 2 miles per week, what is the probability of getting a random sample of size 9 with a mean greater than or equal to 25?

a.) about 50%
b.) almost certain
c.) almost never
d.) impossible, since you would never catch them.

Question # 2 5.0 Points

A simple random sample is one where

a) you decide on a sample size and sample proportionately from the strata.
b) you choose each item with no regard to previous choices.
c) each item in the population has an equal chance of being chosen.
d) all of the above are true.
e) none of the above are true.

Question # 3 5.0 Points

COMPUTE THE MEAN, MEDIAN, AND MODE OF THE FOLLOWING SCORES:
2,8,4,6,5,5

Question # 4 5.0 Points

The median of the 7 numbers {9, 1, 3, 0, 0, -9, 10} is

1) equal to the mode
2) equal to the mean
3) 0
4) 3
5) none of the above

Question # 5 10.0 Points

For questions A to D, use the following data set: 7,3,6,5,1,2,6,6,7,2,6,3,4.

a. What is the mean for this data set?
b. What is the median for this data set?
c. What is the mode for this data set?
d. What is the range for this data set?

Question # 6 10.0 Points

Find the mode for the following data set:
Ford, Chevy, Dodge, Ford, Datsun, Datsun, Ford, Chevy, Dodge.

Question # 7   10.0 Points

A floor manager of a large department store is studying habits of their customers. Suppose he has good reason to believe that an estimate of $600 for the population mean given to him is wrong. He makes preparation to draw a sample but lacks the funds to draw N = 100 as he had planned. How large a sample need he draw in order to estimate the population mean within $100 of the true value with a probability of 0.95? (Assume SIGMA = $500)

Question # 8   10.0 Points

The following are the reaction times in seconds of five people to a particular stimulus: 7,8,6,10,9

a. Determine a 0.90 confidence interval for the mean reaction time for all people to this stimulus.

Question # 9   10.0 Points

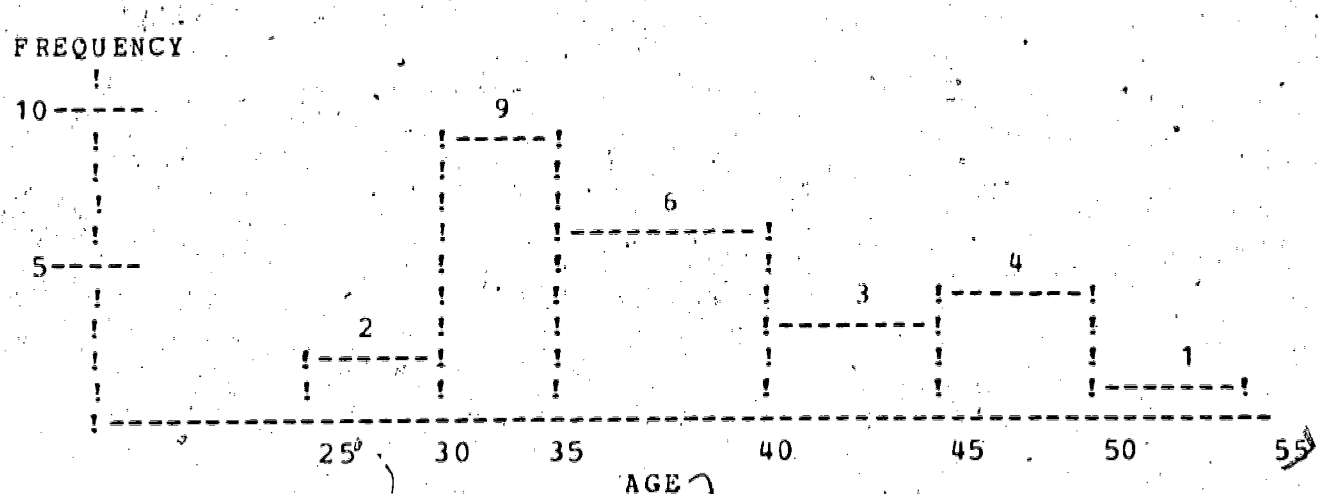
TRUE OR FALSE? IF FALSE, WHY?
AS A FIRST STEP IN SAMPLING VOTOR PREFERENCE FOR PRESIDENT THROUGHOUT THE U.S., A SIMPLE RANDOM SAMPLING PROCEDURE SHOULD GIVE MORE RELIABLE RESULTS THAN A STRATIFIED SAMPLING PROCEDURE WITH THE SAME SAMPLE SIZE.

Question # 10   10.0 Points

A manufacturer wishes to determine the average weight of a certain type of product in order to design the proper package. What size sample is required so that the risk of exceeding an error of .20 pounds or more is .010? Assume that sigma is 1.10 pounds.

Question # 11   10.0 Points

EACH 4TH GRADER IN A SACRAMENTO ELEMENTARY SCHOOL WAS ASKED THE AGE OF HIS FATHER. THE RESPONSES ARE CLASSIFIED IN THE FOLLOWING HISTOGRAM.



THE MEAN XBAR OF THIS CLASSIFIED DATA IS:

A) 35   B) 36.25   C) 37.5   D) 37.7   E) NONE OF THESE

Question # 12 10.0 Points
TRUE OR FALSE. IF FALSE EXPLAIN WHY.

If the results of an investigation show that one decongestant works better than another, there is no need to know the level of significance.

Question # 1 5.0 Points
c.) almost never

MU = 10
n = 9
SIGMA(xbar) = 2

Z = (25 - 10)/ 2
= 7.5

Area beyond (Z = 7.5) = approx. zero

Question # 2 5.0 Points
c) each item in the population has an equal chance of being chosen.

Question # 3 5.0 Points
NONE PROVIDED

Question # 4 5.0 Points
(5): mode = 0
mean = 2
median = 1

Question # 5 10.0 Points
a. 4.46
b. 5
c. 6
d. 6

Question # 6 10.0 Points
Answer - Ford; the mode is the most frequently occurring value.

Question # 7 10.0 Points
N = [Z(ALPHA/2) * SIGMA/e]**2; e = tolerable error
= [(1.96) * 500/100]**2
= 96

Question # 8 10.0 Points
a. XBAR = 40/5 = 8
S(X)**2 = [(-1)**2 + (0)**2 + (2)**2 + (1)**2]/4
= 10/4
= 2.5
C.I. = XBAR +/- t(ALPHA = .10)*S(XBAR)
= 8 +/- (2.13 * .707)
= 8 +/- 1.51

6.49 < MU < 9.51

Question # 9   10.0 Points

NONE PROVIDED

Question # 10   10.0 Points

Using Z = (XBAR - mu)/(sigma/SQRT n)

we get n = ((Z) (sigma)/(XBAR - mu))**2

n = ((2.576) (1.1)/.2) **2

n = 200.73

Therefore, a sample size of 201 is required.

Question # 11   10.0 Points

D

Question # 12   10.0 Points

False, any comparison of differences between two treatments will be significant at some specified level. The real issue is how significant are the results. A statement such as "using a confidence level of 95%, Brand A was found to be significantly better than Brand B at providing relief" clearly indicates the significance of the result and implies that such a result could occur only one time in 20 if the two brands were really equivalent. The statement "using a confidence level of 50%, Brand A was found to be significantly better than Brand B at providing relief" is not nearly as persuasive. It indicates much less confidence in the conclusion in the sense that if Brands A and B were equivalent, differences such as that observed would occur 10 times in 20, or 50% of the time.

This is a Cross-reference between the question numbers on this exam and the question numbers in the File

1-	54	2-	7	3-	16	4-	17	5-	19
6-	62	7-	35	8-	46	9-	10	10-	34
11-	18	12-	51						

How to Lie with Statistics
Spring, 1978
Exam - 3

Question # 1 5.0 Points

If an investigator wishes to decrease the size of his sample which of the following can he do?

a.) increase the tolerable error
b.) decrease the population variance
c.) decrease the tolerable error

Question # 2 5.0 Points

TRUE OR FALSE? IF FALSE, CORRECT IT.

No matter how convincing it may seem one cannot safely rely on data unless you know how it was gathered.

Question # 3 5.0 Points

For a certain normally distributed population, the value of the standard deviation is known, but the value of the mean is unknown. What will be the effects of changes in sample size and in the confidence coefficient on the length of the confidence interval estimate of the population mean?

a. Increasing sample size increases the length, given a fixed coefficient.
b. Increasing the confidence coefficient decreases the length given a fixed sample size.
c. Increasing sample size decreases the length, given a fixed coefficient.
d. None of the above

Question # 4 5.0 Points

Find the mode for the following data set:
Ford, Chevy, Dodge, Ford, Datsun, Datsun, Ford, Chevy, Dodge.

Question # 5 10.0 Points

Consider this class as a sample from a population taking this exam. "Half of the grades for this class will be below average."

a. What estimator is referred to in the above statement?
b. If we assume that grades will have a symmetric distribution, what additional estimators might be used for the same purpose?
c. Why is the mean often not used to indicate average income?

Question # 6 10.0 Points

A caller on a radio talk show reports having surveyed voters in his neighborhood and finding:
66 2/3% favor candidate A.
33 1/3% favor candidate B.
33 1/3% more voters favor candidate A, so he is expected to carry the neighborhood.
Do you accept this conclusion? Why?

Question # 7 10.0 Points

Define the following term and give an example of its use. Your example should not be one given in class or in a handout.

Standard error of a mean

Question # 8 10.0 Points

Define the term "stratified sample" and explain why it would be useful in the following situation. A company is composed of many small plants located throughout the United States. A Vice President of the company wants to determine the opinions of the employees on the vacation policy.

Question # 9 10.0 Points
HEADLINE: MPG for Gas Guzzler skyrockets over
MPG for Econ Scooter!

Data: in miles per gallon

	Gas Guzzler	Econ Scooter
1964	4	25
1968	5	30
1972	8	35
1976	16	40

```
            100 !                          G
 Percent        !
 Increase       !
 Over        75 !
 Prior          !
 Time           !          G
 Period      50 !
                !
                !
             25 !    G
                !    E
                !          E        E
                ------------------------------------------
                     1968      1972      1976
```

(To complete the graph connect the three G points with straight lines to relate the performance of Gas Guzzler. Similarly, connect the three E points to show the trend for Econ Scooter.)

Even though the above graph is correct, explain how it has led to the misleading headline.

Question # 10 10.0 Points
IDENTIFY THE ERRORS IN THE GRAPH

Question # 11 10.0 Points
In statistics we speak often of a random sample. What is a random sample, and why is randomization so important in statistics?

Question # 12   10.0 Points

A teacher is conducting two sections of the same course. As the teacher experiences the two sections it is his/her belief that the students in section one are not doing as well as the students in section two. As a preliminary test of whether he/she should approach the two sections differently the teacher administers a quiz with 20 questions on it to both sections on the same day. Here are the scores for both sections.

Section 1: 20, 19, 18, 18, 17, 16, 16, 16, 15, 13
Section 2: 19, 19, 19, 17, 16, 16, 15, 15, 14, 14

Should the teacher consider the sections as different or about the same? Why?

Question # 1 5.0 Points

a.) increase the tolerable error

Increasing the amount of error which will be tolerated allows for less precision in the results, thereby requiring a smaller sample size. Decreasing the tolerable error would have the opposite effect; and obviously the investigator has no control over the variance in the population.

Question # 2 5.0 Points

True.

Question # 3 5.0 Points

c. Increasing sample size decreases the length, given a fixed coefficient.

Question # 4 5.0 Points

Answer - Ford; the mode is the most frequently occurring value.

Question # 5 10.0 Points

a. The mean if you use the term "average" loosely; more properly, it is the median, since by definition it separates a group of scores into two equal-sized groups of values.

b. The mode and mean.

c. The mean is not often used to indicate the average income because the distribution of incomes like many other variables, tend to be skewed to the right. This results from the fact that most people have low or middle incomes while a small percentage have very high incomes. Therefore, more people will be closer to the minimum than the maximum. The mean is highly influenced by this small number of extremely large values, hence, the median income is commonly reported.

Question # 6 10.0 Points

No, I don't accept it because I have no idea what size sample was used, and I might be suspicious enough to think that it was 3.

Question # 7 10.0 Points

Definition: A measure of variation among values for sample means for a particular sample size. The usual method used to calculate a standard error of a mean for sample size n is to obtain a variance for individual sample elements (S**2), convert it to a variance of a mean for the desired sample size (Divide S**2 by n), and take the square root of that variance. Notice that the same value for S**2 can be used to calculate

standard errors of a mean for several different sample sizes.

Example: Suppose that a random sample of size 10 has provided a variance of 25 with 9 df. That variance of 25 provides the basis for calculating

A: Standard error of a mean for a sample size of 4 = 2.5

B: Standard error of a mean for a sample size of 16 = 1.25

C: Standard error of a mean for a sample size of 25 = 1.00

Symbols: S(YBAR), S(XBAR), SIGMA(YBAR), SIGMA(XBAR)....

Question # 8 10.0 Points

Answer - A stratified sample is one which has been obtained by a procedure in which the frame is divided into non overlapping categories (strata). Sampling units are then selected at random from each stratum thus assuring that all strata are represented in the sample.

For the given problem, I would suggest a type of stratified sampling procedure. Specifically, I would recommend that each plant be considered a stratum and a random sample obtained within each stratum to insure that all plants are represented in the sample. I would further suggest that the sample from each stratum represent the proportional size of that stratum. For example, if plant A employs 25% of the total company's employees, then the sample from plant A should represent 25% of the total sample obtained.

Question # 9 10.0 Points

Answer - The headline is misleading in the sense that it implies mpg is being compared for the two vehicles. Only upon inspection of the data can one see that Econ Scooters have a substantially higher mpg, while their rate of increasing mpg has not been as great. The graph accurately indicates the rate of increase in mpg, but the headline is comparing actual mpg, which is quite different.

Question # 10 10.0 Points

1. LINE OUTSIDE AXIS
2. SLOPPY LABEL (ESPECIALLY VERTICAL AXIS)
3. NO TITLE
4. POOR LEGEND

Question # 11 10.0 Points

A random sample occurs when every observation in the population has a known (usually equal) chance of becoming part of the sample. Randomization is important because the probabilities of samples selected in a random manner are known, thus enabling one to make valid inferences about the population from which the sample was drawn.

Question # 12 10.0 Points

X indicates Section 1:

sum (X(i)) = 168

sum (X(i))/n = XBAR = 16.8

S(X)**2 = 4.1778

S(X) = 2.0440

Y indicates Section 2:

sum (Y(i)) = 164

sum (Y(i))/n = 16.4

S(Y)**2 = 4.0444

S(Y) = 2.0111

Assuming homogeneity of variance and since samples are the same size there is no need to pool. We get:

```
S(XBAR - YBAR)**2 = S(X)**2/n + S(Y)**2/n
                  = 4.1778/10 + 4.0444/10
                  = .822
```

Using a two tail t-test with alpha = .05 and df = 18, we test:

```
H(0):  mu(X) - mu(Y) = 0
H(1):  mu(X) - mu(Y) =/= 0
```

t = (16.8 - 16.4) - 0/SQRT(.822) = .4/.907 = .4411

t(critical) = +/- 2.101

Therefore continue H(0) or consider the sections to be about the same.

This is a Cross-reference between the question numbers on this exam and the question numbers in the File

1-	57	2-	1	3-	43	4-	63	5-	13
6-	24	7-	30	8-	60	9-	64	10-	21
11-	11	12-	42						